

Thomas Bennett
CS 4395.001 - UTD Spring '23
Portfolio Assignment 1: Overview

This program fixes the formatting of a csv file. Each line in the file is formatted and then saved as a Person object. A dictionary of Person objects is then created and saved using the Python library, pickle.

To run the program, download the entire Homework 1 directory. Navigate to the location of the program file "HW1_trb090020.py" and open a terminal or command shell. Run the program with the relative path to the data.csv file. Some users may need to use "python" or "python3" to invoke Python.

Example command:

```
"python3 HW1_trb090020.py ./data/data.csv"
```

Alternatively, use your favorite Python IDE. I used PyCharm to develop the program. You can include arguments to the program in PyCharm by going to Run > Edit Configurations. The Parameters field allows you to provide parameters to the run of the program.

Python makes file I/O very simple and has lots of handy built-in classes and methods for all the tasks required of this assignment. However, the object-based nature of Python can lead to confusion. It could be easy to confuse regular expression objects with regular expression pattern strings. Since Python allows regular expressions to be used either way (strings or compiled objects) it is probably best to pick one way and stick with it.

I have used Python in previous classes so most of this assignment was somewhat of a review. Big Data Management & Analytics uses Python with Spark. I did learn a little bit about regular expressions in Python. I thought about implementing this program using Spark and dataframes, but I decided against it.